Specimen Databases: A Case Study in Entomology using Web-based Software

Randall T. Schuh, Sheridan Hewson-Smith, and John S. Ascher

The earliest versions of biological specimen databases were probably card files and ledgers in collections of vertebrate specimens. The sophistication and value of such databases has grown over time because of the increasing accessibility and power of computer technology, the accumulated number of digitized specimen records, and the application of specimen information to problems in land management, climate change, plant–insect associations, epidemiology, and range modeling. Although relevant technology has existed for many years and is continuing to improve, design and implementation of specimen databases is hindered by the complexity of the decision-making process.

The structure of the data associated with biological specimens is well understood and has been for some time (TDWG 2010). It is not surprising that most of the existing databases have adopted similar data models for taxonomy, geographical information, and specimen data. Entomological collections, however, are distinctive because of the large numbers of taxa, very large numbers of specimens, the pin mounting of most taxa, fragility of the specimens, use of miniaturized labels, and longstanding absence of unique specimen identifiers.

Insects and other terrestrial arthropods have the potential to reveal a high level of detail for patterns in nature simply because of their numbers. They do great harm to crops, forests, and human health, but also provide great benefits as pollinators and natural enemies, which emphasizes the value of ready access to collection-based information. But, if entomological data are to be used to their full potential and achieve a status for insects that is similar to vertebrates and higher plants in the information age, entomological databases must be tailored to their particular attributes. In this article, we describe the design and development of a Web-accessible specimen database, discuss the potential use of this system across projects and institutions, and explain the features of this database that will extend the useful life of the software application well into the future.

Background: The Plant Bug Planetary Biodiversity Inventory Project

In 2003, the National Science Foundation (NSF) began funding Planetary Biodiversity Inventory (PBI) projects in an effort to ramp up the study of biodiversity worldwide. The Plant Bug PBI project began in September 2003 with an award to the American Museum of Natural History (AMNH). As the principal investigators on the award, Randall Schuh (American Museum of Natural History, New York) and Gerasimos Cassis (University of New South Wales, Sydney) assembled a team of five senior scientists to work in conjunction with four postdoctoral researchers, three Ph.D. students, several undergraduate students, and multiple technical support personnel, including co-author Sheridan Hewson-Smith.

The objective was to assemble an international treatment of the monophyletic group comprising the plant bug subfamilies Orthotylinae and Phylinae (Heteroptera: Miridae). Countries of particular interest because of their high species richness but incomplete knowledge included Australia, South Africa, and Mexico. At least 1,000 undescribed species were predicted to exist worldwide, beyond the 4,000 species already described. More information can be found at the project Web site, http://research.amnh.org/pbi/.

The PBI research team was located on three continents, which presented a set of challenges not previously addressed by the entomological community in the Internet Age. Some content relevant to the project was already available on the Web, in particular the *Systematic Catalog of Miridae* (http://research.amnh.org/pbi/catalog). Nonetheless, the original proposal to the NSF contained only a vague description of specimen data capture and dissemination, with no clear vision as to how this pillar of the project would function—an indication of the general state of affairs as of 2002.

We believe the design and development of the Plant Bug PBI database may provide valuable lessons for others who might venture into database development and implementation. The subject is particularly relevant at a time when discussions are taking place within the NSF, at the highest levels of the U.S. government (Interagency Working Group 2009), and elsewhere about how to make greater amounts of specimen data available and develop the funding to do so.

In addition to integrating the data capture activities of investigators in Canada, the United States, Russia, and Australia, the PBI team was faced with delivering that data for use by its members, research scientists, and other interested parties. As the PBI project progressed, we were approached by investigators working on bees, scorpions, spiders, wasps, and other organisms, all of whom had similar specimen database needs that could be accommodated by

the PBI database. Thus, our story is about a system that is successful beyond the limits of the Plant Bug PBI project and therefore is of potential interest to the broader entomological community.

Our approach to decision-making included multiple factors. We had to consider tradeoffs between complexity, functionality, and cost, and the effect these decisions would have on a team-based research effort.

Database Environments and Sources

In 2003, numerous database applications were available. Our first task was to determine which existing database software would best achieve our objectives. To answer this question, we first chose among three options, which were not mutually exclusive:

1. Flat-file vs. SQL-compliant relational databases
2. Proprietary vs. open-source software
3. Stand-alone vs. LocalAreaNetwork (LAN)-based vs. Web-based applications

To better assess our options, we acquired copies of BioLink (created by *Commonwealth Scientific and Industrial Research (Organisation ... Australia,)* in 1999)[1] and BIOTA (Colwell 2009) software, each of which function as stand-alone applications. We visited an installation of the NSF-funded, LAN-based application SPECIFY (2010) and asked many questions about its functionality and potential suitability for our project. We examined approaches being used by the Australian Museum in Sydney (then the home of coPI Cassis), talked with knowledgeable colleagues, and evaluated systems used by other NSF-funded taxon-based projects, such as those in the PEET program (Partnerships for Enhancing Expertise in Taxonomy).

Choice 1: Ease of implementation and availability has led many people to adopt spread-sheet (flat-file) database software. This method is fast and can be used by almost anyone, working with programs such as Microsoft Excel, because there are few technical hurdles to surmount. Its weakness is that the user has limited control of the structure and accuracy of the data during the data entry process because all of the information must be reentered for every database record. This leads to errors, and in the worst cases, many unique errors. There also are conspicuous limitations on the ultimate capacity of such systems in terms of field size and the numbers of columns and rows.

A readily available alternative is a relational database that conforms to SQL (structured query language) standards (http://en.wikipedia.org/wiki/SQL). This kind of software does not require redundant data entry. It provides greater data quality control and allows selected data to be exported easily in a flat-file format (such as a spreadsheet). It is also ready-made for creating hierarchic data structures, such as biological nomenclature. Nonetheless, history has not always been kind to specimen database developers, in part because not all relational databases were SQL compliant; many did not allow for easy transfer from one software platform to another (e.g., earlier versions of FileMaker), or the software could not handle a sufficiently large number of records (Microsoft Access). After considering these issues, we chose SQL-compliant relational database software that had no practical limitation on field size, table number, or record number.

[1] For more information see http://www.its.csiro.au/news/mediarel/mr1999/mr99116.html.

Choice 2: Members of our team knew the realities of using commercial software: new releases of programs over time render older versions obsolete and require costly upgrades. Our awareness of software licensing issues and planned obsolescence led us to learn more about "open-source" software: program development is not driven by monetary reward, the path of migration can be done at little or no cost for the underlying software, and new versions are not necessarily designed to make prior versions obsolete.

Choice 3: Our decision-making about networking was dictated by the capabilities that would be needed by our project team. We could adopt a flexible stand-alone product such as BioLink, but we would then need to devise a method for amalgamating the data entry efforts of individual participants. Using a LAN-based approach would have provided greater within-institution capability, but would still have required amalgamation of data across institutions. Either of these options allowed for the adoption of existing software products, but still posed substantial technical and logistical problems. Alternatively, a Web-based approach would allow participants anywhere in the world to enter data consistently through a uniform portal to a single server. The biggest drawback would be the need to write the entire application from scratch because no suitable products were available at the time. Despite the challenges presented by software-application development, we chose the Web-based alternative because of its fundamental benefits.

- We would largely (if not completely) avoid the need to amalgamate and error-check data from individual participants, or to rationalize discrepancies across various data entry efforts.
- We would have to maintain just one application on one set of servers.
- Tasks such as georeferencing (discussed in more detail below) could be centralized to avoid redundancy of effort and therefore benefit all database users.

Biological Specimen Data Structure

After years of discussion, the collections community has largely agreed on the innate nature of natural history specimen data and the way those data can best be structured in a relational format (TDWG 2010). We therefore adopted, with little modification, the table and field structure being used by the Australian Museum in Sydney. This data format is essentially the same as that in BioLink, BIOTA, and SPECIFY, among other databases. The general nature of that structure can be seen in Fig. 1.

Data Standards and Protocols

Much effort has been expended on the question of data standards. The norms in the field have been established by the Taxonomic Database Working Group (TDWG), the Natural Science Collections Alliance (NSCA) (TDWG 2010), and the Global Biodiversity Information Facility (GBIF 2010). Standards include Darwin Core and Darwin Core II; protocols include DiGIR (Distributed Generic Information Retrieval) and others. We laud these efforts and have established a system that is compliant with them. Many PBI fields anticipate Darwin Core II fields, a good illustration of how non-standard fields of proven utility will eventually become standard fields. Given this historical reality, it makes little sense to wait until a field is added to the standard before capturing relevant data. In addition to compatibility with these standards and protocols, we believe the decision-making rules outlined in this article also require careful consideration.

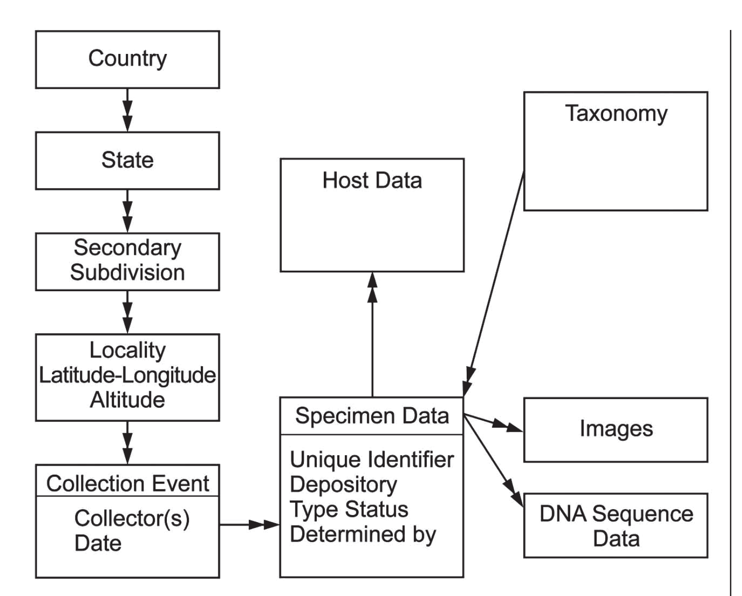


Fig. 1. Simplified version of schema used in PBI database. Double-headed arrows indicate the many (child) side of one-to-many (parent–child) relationships.

Approaches to Data Capture

Following certain design fundamentals, such as isolating the activities of writing to, editing, and reading from the database, will result in a more robust product. Although such considerations may seem mundane, it is important to realize that even though a single interface (such as a spreadsheet) can serve both write and read functions, it will not do either in an optimal way. Thus, the design of the user interface is critical. If the interface works well, all other user functions should follow with comparative ease. For this reason, the PBI team put much of its effort into designing and implementing the user interface. If the PBI project were to be a success, the data capture process would have to be rapid and accurate because we predicted the need to capture ~500,000 specimen records in 5–6 years. Simply put, we had to design software tailored to the task at hand, verify that our Internet connections had sufficient capacity, and end up with a product that all of our members could master quickly and use in their daily work.

Forms. A typical and powerful approach for entry of information into a relational database is a form. Forms map boxes shown on the computer screen to data fields in various tables in the database. The form controls the mapping function; it also provides a substantial amount of control over and feedback about the data that are being entered, thereby improving the accuracy of the recorded data. Some of these same functions can be controlled at the table level, where the data are actually stored in the database. Using a combination of these two mechanisms, a broad range of activities can be controlled, in terms of restriction and permission. Examples include data types, which can control whether a data field accepts numbers, letters, or both; how a date field stores data and in what format those data are displayed to the user; and whether the initial letter in a field should be capitalized, as in the case of generic names. The possibilities are almost endless.

Such restrictions on data format might be seen as a drawback by some users because specimen label data frequently come in quirky forms or are incomplete. Nonetheless, in conjunction with data accuracy, uniformity of structure makes the data valuable over time. Using a structured database may require some compromise when compared with entering plain text, but it ultimately achieves efficient work flow, high-quality data entry, and publication-quality data ready for immediate use.

Verbatim Data. The PBI project chose to accept a structured data design and then write label data to it, instead of transcribing the data verbatim. Some label data will be transformed from their original form during the data transcription phase. Although some users see this as a drawback, we believe benefits are to be gained from the direct recording of highly structured label data. Such data can be searched on a uniform set of criteria, making it more straightforwardly useful. Also, incorporating unique specimen identifiers (see below) unequivocally connects almost every record in the database to an individual specimen. Therefore, the original data are always retrievable, even though answering some historical questions may require recourse to the specimen(s) from which the data were captured. Although it is not our practice to routinely capture verbatim data, the PBI database allows for the creation of "notes" on the data, including cases where such annotation is desirable. Finally, capturing data in a structured format renders them immediately publishable, whereas verbatim data must be transformed and checked before they can be published. Our approach does not preclude eventual capture of verbatim data because it is possible to associate images of labels (or any other images; see below) with previously captured specimen records.

The PBI Database from the Users' Perspective

After we considered the database needs and resources available, we designed the PBI database with five main modes:

1. Log-in Screen. The log-in screen is accessible through any Internet connection. It provides secure access to the database, the choice of the taxon-specific system (user profile), and mode (write, edit, report).

2. Data Entry (Museum) Mode. The Museum Mode interface is divided into taxon information, locality information, collection event information, specimen information, and host plant information (Fig. 2). To ensure a sufficiently complete data set, no specimen information can be recorded without essential taxon, locality, and collection event information.

The screen design is intended to expedite the data entry process by the following means: all fields involved in data entry for a typical specimen are visible on a 1024 × 768 or higher resolution screen without scrolling, even though multiple tables are presented in a single form; the thematic layout of the screen orients the user during the process of specimen data capture; navigation through the screen can be accomplished by tabbing, which minimizes the use of the mouse; and any one of the five panels on the screen can be cleared without affecting the remaining panels, or the entire screen can be cleared.

These design elements assist in the logical flow of work. For example, if data are being entered for all specimens for a given species, the information in the taxonomy panel can remain static, while information in the remaining panels changes. If multiple species are being entered from a single collection event, information in the locality and collection event panels can remain static, while information in the taxonomy and specimen data panels changes.

Fields that offer choices, such as family, genus, country, collector, and depository, allow the selection to be made simply by typing and do not ordinarily require the use of the mouse. Entry of the first few letters in a word will usually bring up the proper choice, which can

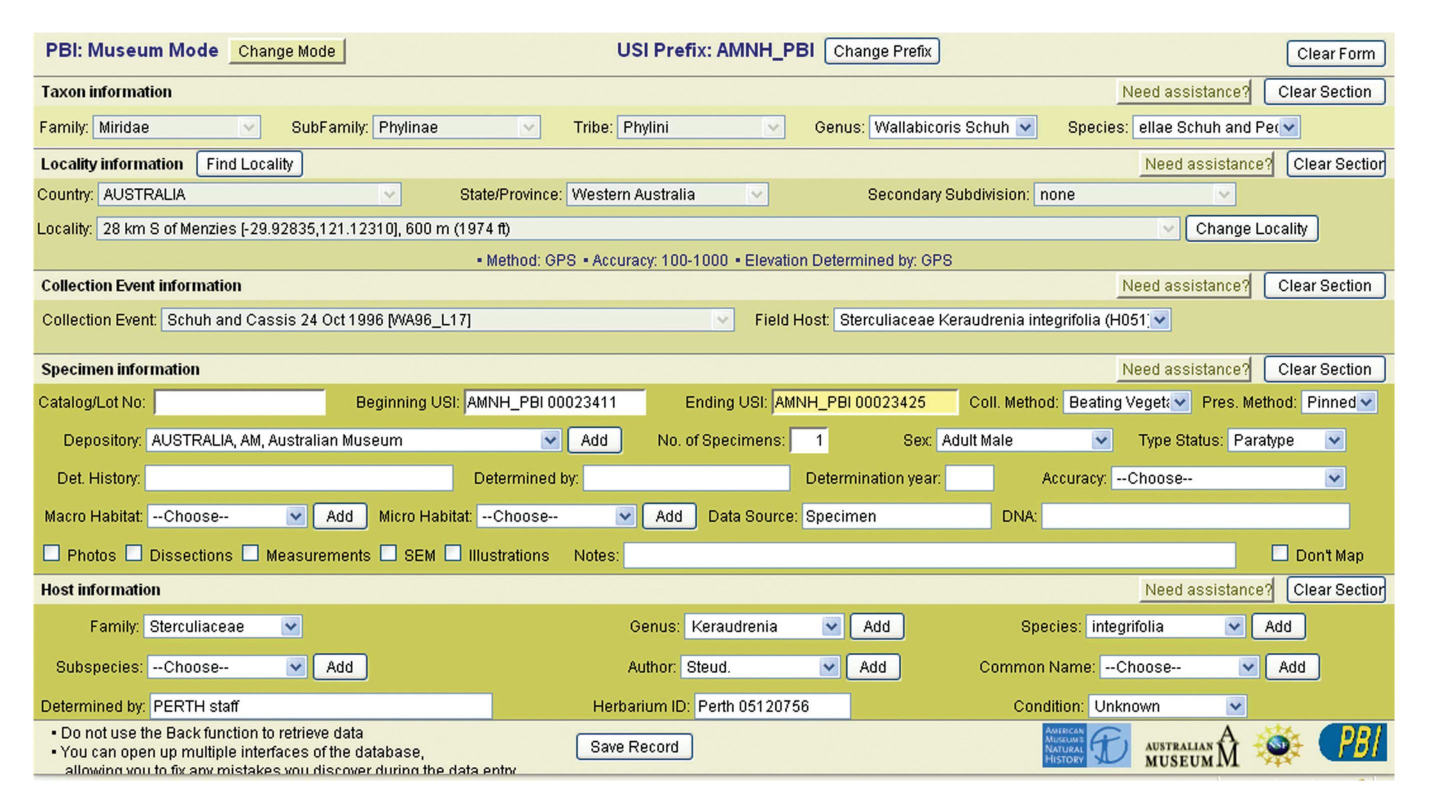


Fig. 2. Data entry (Museum Mode) screen from Plant Bug System of the PBI database. Notice division of form, from top to bottom, into five sections, each representing a distinct aspect of the data, and each subject to independent control during the data entry process. Note the "Clear Form" and "Clear Section" buttons on the right side of the screen. The "Need assistance" buttons provide pop-up help screens.

then be accepted by hitting the enter key. The position of the cursor is always obvious, the functionality of the drop-down fields is always the same, and navigation from panel to panel is seamless.

Features such as these have engendered acceptance of our database by a diverse group of users. Much of this functionality is controlled by the browser (Mozilla, Internet Explorer, etc.), the choice of which can greatly facilitate, or hinder, the data entry process. It is our experience that the Mozilla Firefox browser works particularly well.

Fields with drop-down functionality do not accept new data directly. The uniformity and integrity of data in these fields (such as country names and plant family names) benefit from the data being entered ahead of time so that all users are presented with uniform information from an authoritative source from which to make choices. This feature increases the accuracy and consistency of the data.

Taxon Information. Taxonomic information is always entered first, after which it can also be retained for as long as data are being entered for specimens of the same taxon. The data are organized hierarchically, as shown in Fig. 1, so that choice of a family produces a list of corresponding subfamilies, and so forth down to the level of species. Only genus and species names can be entered directly through Museum Mode.

Locality Information. Localities can be looked up using the "Find Locality" button and searching for a string of characters in the Locality field, or they can be retrieved by entering data in the series of hierarchically organized fields country, state/province, and secondary subdivision. Locality data are shared by all users of the database, no matter what system they are using (e.g., Plant Bugs, Bees, Spiders); this helps to limit redundancy and promote efficiency by pooling georeferencing effort. Latitude/longitude data can be entered in any format and will always be returned in degrees and decimal parts thereof, which facilitates mapping. Latitude/longitude and elevation data are accompanied by metadata fields that indicate the source of the data and the confidence interval.

Collection Event Information. In the PBI relational model (Fig. 1), the table containing information on dates and collectors forms a child relationship with the locality table, such that a given locality may have many collection events attached to it. These data will be displayed as a drop-down list once a locality with associated collection events has been chosen. A form is also available to enter new collection event data from scratch. The date format was chosen to avoid ambiguity arising from different conventions for recording dates in numerical form, while at the same time being readily utilizable by an international team. All records must have at least a start date.

Specimen Information. All specimen-specific data are recorded in this section, and the connection is made to a machine-readable code: the unique specimen identifier (USI). Two aspects of this specimen information panel deserve mention. First, USI codes can be entered as a series, whether manually or with an electronic scanner; e.g., AMNH_PBI 00000020—AMNH_PBI 00000033 (beginning USI—ending USI). Thus, records for a group of specimens with otherwise identical data can be entered in a single action. This function expedites data entry for a series of specimens. Second, this panel includes an identification history, which allows for capture of information on the changing identity of specimens over time, including taxon name, author, identifier, and year.

Host Information. The PBI project focused on a group of plant bugs, phytophagous insects that show substantial host specificity; therefore, the capture of host plant data was a central part of our database concept. These data are captured in the bottom panel on the main data entry (Museum Mode) screen (Fig. 2). They allow for the inclusion of voucher specimens and images. As with arthropod names, error-checking against available authority files (such as those in the Integrated Taxonomic Information System [ITIS 2009]) can

greatly improve the accuracy and consistency of botanical nomenclature as rendered by entomologists and their technicians. Technology to perform such checking was designed by John Pickering of the Polistes Foundation and is available through the Web aggregator DiscoverLife.org.[2]

Many entomologists can answer the question, "What are the host plants of insect species A?" It is much harder for them to answer the question, "What are the predators, pollinators, or other visitors of plant species B?" In the case of phytophagous insects, parasitoid Hymenoptera, and others, assembling information that can be cross-referenced can be a daunting task. Once our databases are sufficiently well populated, we will have much stronger tools available for addressing these questions.

3. Edit Mode. All data in the PBI database can be accessed through a series of "edit screens" (forms). Edit screens organize data in various ways, ranging from a simple list to more complex displays where data for all specimens of a given species are presented and can be chosen for individual or mass editing. In the Edit Mode, rarely entered data such as family-group names can be added to the database.

Editing the data, particularly en masse, is something that should be done with caution. For this reason the edit screens always ask for confirmation of intended changes before those changes are written to the database. Safeguards are programmed to prevent the accidental writing or alteration of thousands of specimen records because of one or more misplaced digits in the USI. All records in the database are stamped with the date of creation and the name of the person who entered or modified the data. This information is revealed on the edit screens and makes it possible to track record creation and modification (Fig. 3).

Multiple views on the database can be open at the same time in separate browser windows. When using this capability, errors or omissions in data can be corrected as soon as they are recognized by toggling from one screen to another. This allows the Museum Mode data entry screen to maintain its information while erroneous entries are simultaneously corrected individually or en masse in Edit Mode. This multitasking approach is made possible by the Web-based nature of the application and the multithread capability of the database.

4. Report Mode. This mode allows the user to query almost all of the data in the database. It evolved over the course of the project as participants determined their research and collection management needs.

Because the database was designed to facilitate the preparation of revisionary studies (e.g., Forero 2008, Tatarnic and Cassis 2008, Platnick and Dupérré 2010, Schuh and Pedraza 2010), one of the first—and most valuable—reports we prepared was for specimen data written as formatted text that can be pasted directly into the manuscript for a revisionary study (Fig. 4). Traditionally such data were typed and retyped, with no way to reuse them. With a database, they can be generated in a prescribed format almost instantaneously; if corrections are needed, they can be made to the source and the report generated once again. On a general level, we provide for queries on a broad range of criteria (Fig. 5). The results can be exported to a spreadsheet or flat-file database in the form of a tab-delimited file. These data can then be manipulated to suit the needs of the individual user.

Locality and host labels can be prepared easily in Report Mode, and specimen labels therefore match the information in the database, no matter when the labels are created. If all locality and collection event data, including habitat and host images, are recorded in the database during fieldwork or immediately upon return from the field, that data can be used to generate labels, will be available for eventual attachment to all specimen records, and will promote production of reliable and uniform specimens-examined reports.

5. Administrative Mode. An administrative module allows for the registration of users and the assignment of usernames and passwords; access to this module is restricted to only those users with administrator privileges.

Enhancing the Value of Specimen Data

Incorporating an Image Library. High-quality digital images have become the norm for biological specimens, preserved and in the field (Fig. 6). For the PBI project, images of interest included

[2] The mission of Discover Life is to "to assemble and share knowledge in order to improve education, health, agriculture, economic development, and conservation throughout the world." This Web aggregator provides a wealth of systematic and other information, including authoritative lists of taxon names, identification aids, and images, on a wide range of organisms.

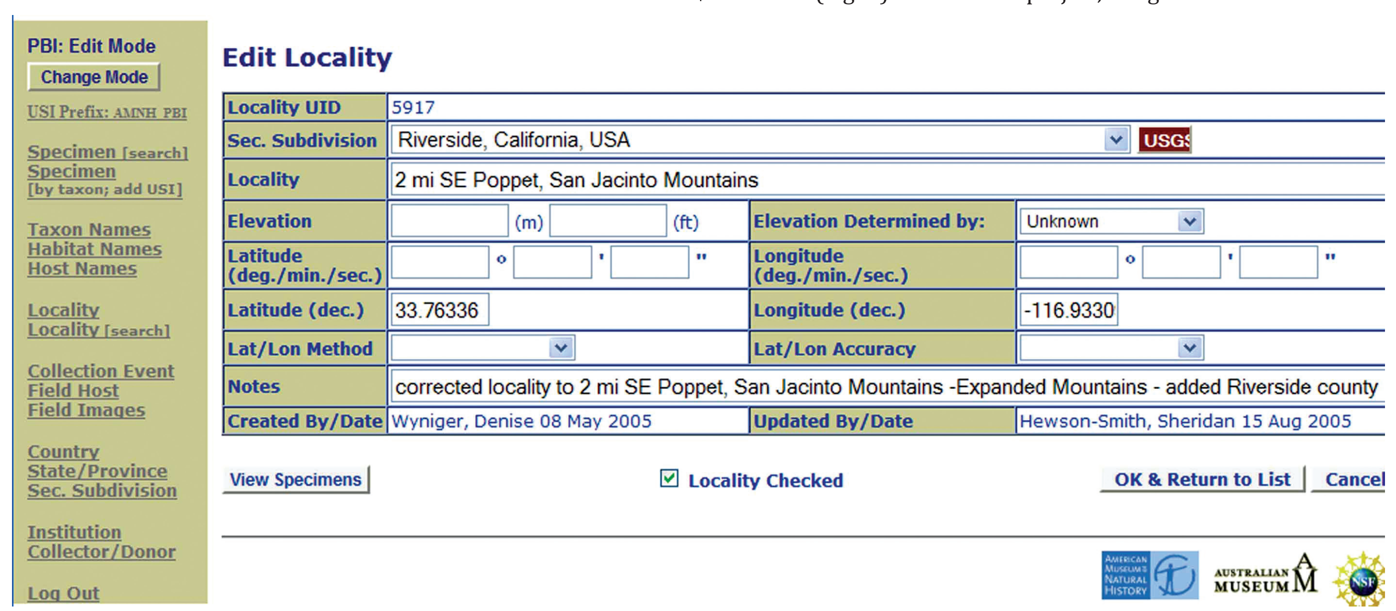
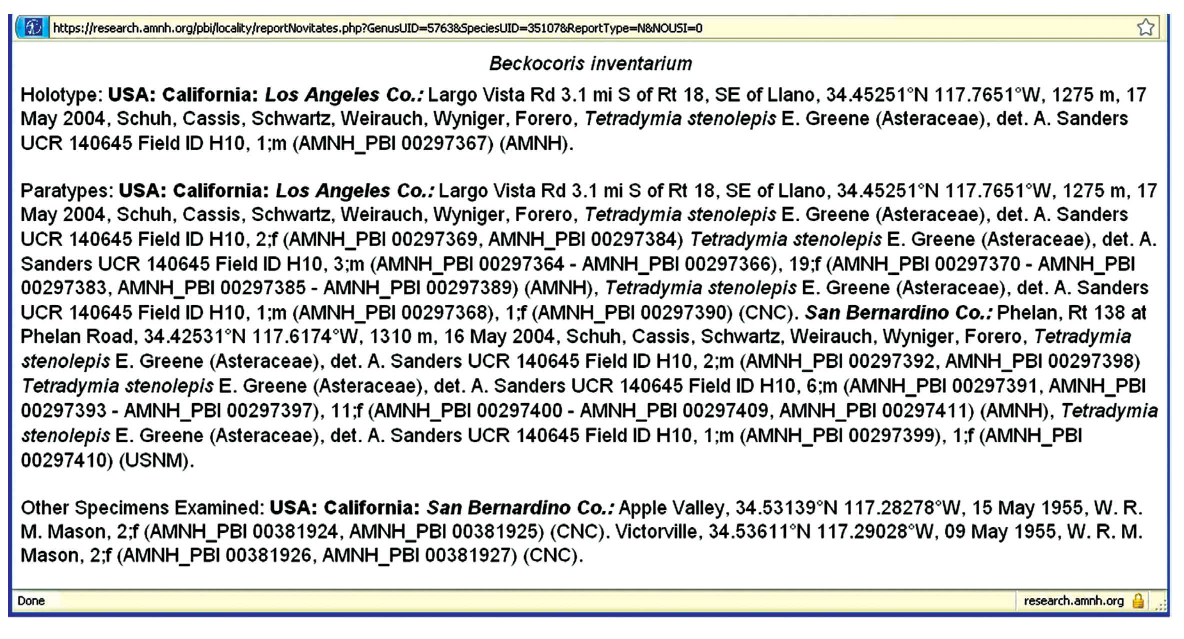


Fig. 3. Locality edit screen showing "Created By/Date" and "Updated By/Date" fields for tracking user information.



https://research.amnh.org/pbi/locality/reportNovitates.php?GenusUID=5763&SpeciesUID=35107&ReportType=N&NOUSI=0

Beckocoris inventarium

Holotype: **USA: California:** ***Los Angeles Co.:*** Largo Vista Rd 3.1 mi S of Rt 18, SE of Llano, 34.45251°N 117.7651°W, 1275 m, 17 May 2004, Schuh, Cassis, Schwartz, Weirauch, Wyniger, Forero, *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 1;m (AMNH_PBI 00297367) (AMNH).

Paratypes: **USA: California:** ***Los Angeles Co.:*** Largo Vista Rd 3.1 mi S of Rt 18, SE of Llano, 34.45251°N 117.7651°W, 1275 m, 17 May 2004, Schuh, Cassis, Schwartz, Weirauch, Wyniger, Forero, *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 2;f (AMNH_PBI 00297369, AMNH_PBI 00297384) *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 3;m (AMNH_PBI 00297364 - AMNH_PBI 00297366), 19;f (AMNH_PBI 00297370 - AMNH_PBI 00297383, AMNH_PBI 00297385 - AMNH_PBI 00297389) (AMNH), *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 1;m (AMNH_PBI 00297368), 1;f (AMNH_PBI 00297390) (CNC). ***San Bernardino Co.:*** Phelan, Rt 138 at Phelan Road, 34.42531°N 117.6174°W, 1310 m, 16 May 2004, Schuh, Cassis, Schwartz, Weirauch, Wyniger, Forero, *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 2;m (AMNH_PBI 00297392, AMNH_PBI 00297398) *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 6;m (AMNH_PBI 00297391, AMNH_PBI 00297393 - AMNH_PBI 00297397), 11;f (AMNH_PBI 00297400 - AMNH_PBI 00297409, AMNH_PBI 00297411) (AMNH), *Tetradymia stenolepis* E. Greene (Asteraceae), det. A. Sanders UCR 140645 Field ID H10, 1;m (AMNH_PBI 00297399), 1;f (AMNH_PBI 00297410) (USNM).

Other Specimens Examined: **USA: California:** ***San Bernardino Co.:*** Apple Valley, 34.53139°N 117.28278°W, 15 May 1955, W. R. M. Mason, 2;f (AMNH_PBI 00381924, AMNH_PBI 00381925) (CNC). Victorville, 34.53611°N 117.29028°W, 09 May 1955, W. R. M. Mason, 2;f (AMNH_PBI 00381926, AMNH_PBI 00381927) (CNC).

Done research.amnh.org

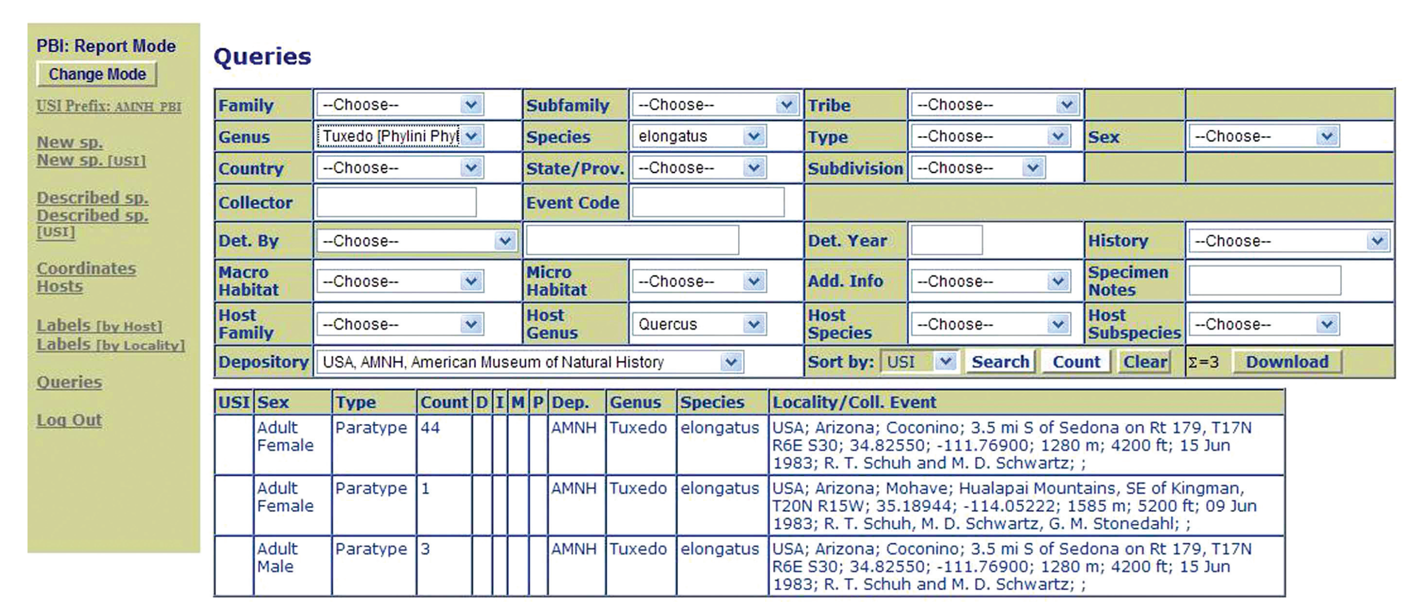
Fig. 4. Specimens-examined output from PBI database formatted for pasting directly into revisionary publications.

habitus and close-up images of taxa, images of host plants, and images of habitats. The PBI database connects all specimen images to their respective specimens by using the unique specimen identifier as the root of the image name. The system enables straightforward association of images with specimen data and allows the retrieval of other data relevant to the image, such as locality information.

Images of hosts and habitats are manually linked to collection event and/or host records in edit mode. The images can be stored as thumbnails, sized for delivery at screen resolution over the Web, or stored at higher resolution for use in publications, posters, and other presentations where high-quality reproduction is required.

Unique Specimen Identifiers. The earliest applications of unique specimen identifiers were probably catalog numbers of the type applied to many vertebrate specimens. The technological solution to this problem has been the use of "barcodes," which was initially implemented at the Costa Rican Instituto de Biodiversidad (InBio). Although incorporating barcoding technology may alter one's work routine, the results benefit revisionary studies, loan management, and the tracking of voucher specimens. Our experience in the use of barcodes shows that machine-reading produces greater accuracy in data recording and data retrieval than when codes are entered by hand.

The PBI project uses matrix code labels, a variant of the barcode, because of their potential for small size and storage of a very large amount of information. The label seen in Fig. 7a measures <5 mm^2 for the matrix code, which can store almost a billion unique numbers per alpha code. For example, the project alpha code "AMNH_PBI" remains static while the numbers change. Other projects using the PBI database have unique alpha codes and also start the numbering from 1; e.g., "AMNH_BEE 00000001." The project code in combination with an eight-digit number allows for a phenomenal range of records that goes beyond the capacity to uniquely identify every specimen in all collections in the world. Use of project alpha codes (acronyms) has an additional value in that it allows institutions and/or projects to be branded—as preferred by collection managers, curators, and deans of collections at institutions or by funding agencies.



USI	Sex	Type	Count	D	I	M	P	Dep.	Genus	Species	Locality/Coll. Event
	Adult Female	Paratype	44					AMNH	Tuxedo	elongatus	USA; Arizona; Coconino; 3.5 mi S of Sedona on Rt 179, T17N R6E S30; 34.82550; -111.76900; 1280 m; 4200 ft; 15 Jun 1983; R. T. Schuh and M. D. Schwartz; ;
	Adult Female	Paratype	1					AMNH	Tuxedo	elongatus	USA; Arizona; Mohave; Hualapai Mountains, SE of Kingman, T20N R15W; 35.18944; -114.05222; 1585 m; 5200 ft; 09 Jun 1983; R. T. Schuh, M. D. Schwartz, G. M. Stonedahl; ;
	Adult Male	Paratype	3					AMNH	Tuxedo	elongatus	USA; Arizona; Coconino; 3.5 mi S of Sedona on Rt 179, T17N R6E S30; 34.82550; -111.76900; 1280 m; 4200 ft; 15 Jun 1983; R. T. Schuh and M. D. Schwartz; ;

Fig 5. A query in Report Mode using multiple criteria. Use of the "Download" button will export a tab-delimited file.



Fig. 6. Example images from the PBI database. Left, specimen images; right, a corresponding host image.

The compact size of these codes makes it possible to incorporate them into the collection management routine of most collections of terrestrial arthropods and with only a modest expansion of space when used in pinned-insect collections (Fig. 7b). Unless obscured by a particularly large insect body, they can be read by machine with the specimen in place in the collection, or, in the absence of a code-reading device, understood from the alpha–numeric version of the code printed on the label. Additional information on the label Fig. 7a identifies the repository of a specimen without accessing the database, allows for precise attribution of specimen ownership (branding), clarifies the physical location of the specimen, and helps in organizing specimens for return of loans.

Georeferencing. Attaching latitude/longitude data to localities increases the value of specimen data because it facilitates such activities as mapping, ecological modeling, and GIS analysis. Latitude/longitude data can be entered in Museum Mode on the "Add Locality" form, or georeferencing can be done subsequently in Edit Mode. Technology and the Internet have greatly enhanced the speed and accuracy of georeferencing. Numerous on-line gazetteers allow for rapid georeferencing of many localities in most countries. Two examples are the Fuzzy Gazetteer (2010) and the Geographic Names Information System (GNIS, 2010). Google Earth also facilitates georeferencing, providing detailed gazetteers and the ability to visualize the point or area being referenced. Automated tools such as *GEOLocate* (Fig. 8; 2010) and *BioGeomancer* (2010) can also facilitate georeferencing through intelligent automation of the process for large blocks of records.

For large numbers of records from historical specimens that did not include latitude/longitude data, the PBI team was able to batch-process the data using *GeoLocate*, make additional checks on their accuracy using other Internet-based gazetteers, and import them to the database. It was also possible to import a large number of records from field work for which latitude/longitude and associated metadata had been captured in the field with a GPS device. After processing as many records as possible using these techniques, we georeferenced newly entered data on a record-by-record basis. Because all of the fields pertinent to georeferencing are stored in the locality table, the process of exporting records for georeferencing and re-import is relatively straightforward.

Finally, we employed a single person to do a large part of our georeferencing for the Plant Bug PBI project. This decision guaranteed consistency and helped to create a feedback loop to continually improve our techniques and capabilities.

Additional error-checking occurs when the DiscoverLife Global Mapper (see Fig. 9) is used to map records. Records with simple errors (e.g., absence of the minus sign preceding the longitude of a record from the Western Hemisphere) are automatically corrected by the DiscoverLife mapping program (also designed by John Pickering of the Polistes Foundation), and all records flagged as erroneous can easily be visualized using a "map inconsistent points" feature. Use of automated error-checkers reduces the problems that arise when georeferencing is done by relatively inexperienced technicians.

For the data to be truly useful, Web-quality and publication-quality maps are necessary. Discover Life and HeteropteraSpeciesPage provide maps of Web quality. We have incorporated a Web-based utility to generate publication-quality maps (http://research.amnh.org/pbi/maps/) without using an independent GIS software package, a feature that works hand-in-hand with our ability to generate publication-ready output for specimens examined.

Maximizing Efficiency of Data Capture

Any approach to specimen databasing should be evaluated under one or more criteria. We might first consider ***efficiency***. This might be measured in terms of numbers of specimen records captured per unit time. It might also be measured by the accuracy of those records. We might also wish to ***minimize specimen handling*** as a way of limiting damage to inherently fragile specimens. This can also be viewed as an aspect of efficiency, but should not be overemphasized. Data capture based on specimen order in the collection may minimize specimen handling at the outset but would certainly not minimize it over the longer run, and is therefore unlikely to be the most efficient approach, because almost every specimen will have to be rehandled before the data-capture process is complete. As another alternative, we might ***prioritize data value***. Under this approach, specimens of a given taxon in poor condition, or those with incomplete or ambiguous data, might be set aside as not yet worthy of the effort to enter their data. ***Novelty of data*** is another consideration, especially with reference to insect collections, where much data associated with already curated specimens may remain unpublished.

Database Specimens Used in Revisions. Revisionary studies produce specimens with the most accurate identifications. For this

(a)

AM_ENT

AMNH_PBI 00388325

(b)



Fig. 7. a.) Machine-readable matrix code label used for unique specimen identification in PBI project. The unique alphanumeric string encoded in the matrix appears on the lower margin of the label. The AM_ENT in the upper left identifies the collection of origin for the specimen. b.) Example usage of matrix code USI labels in the American Museum of Natural History pinned-insect collection.

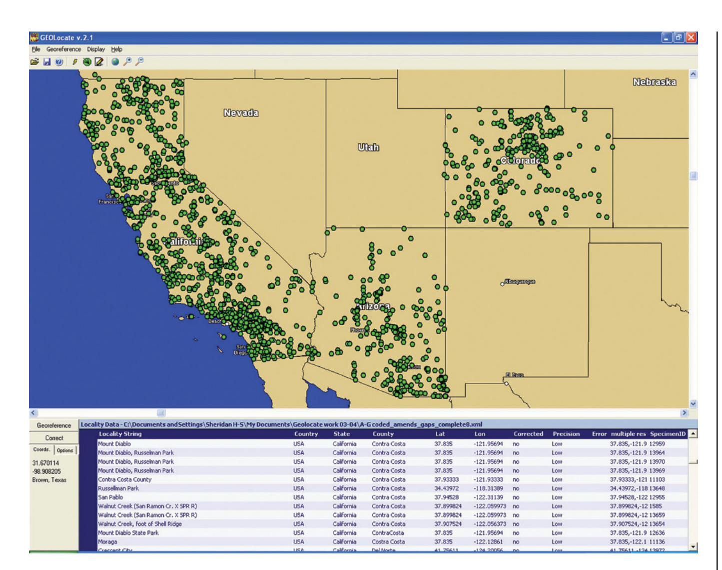


Fig. 8. Screen-shot from GeoLocate showing the feature that allows automated georeferencing of multiple localities.

reason, these specimens offer an obvious starting point for data capture under the above criteria. If data capture is done as part of the revisionary process, the results can be incorporated directly into the study. If data entry is done after the revision, the accuracy of identification will be maintained, although updates of nomenclature may be needed. The data capture in this approach is limited because large amounts of specimen data in a collection will not be recorded because those specimens have never been used in revisionary studies. If the goal of the project is to capture truly novel data, specimen records already published in revisions might be given lower priority, despite the above considerations.

Database Specimens from the Field. When database records are created as material comes from the field (i.e., proactive databasing), the level of taxonomic identification is often too general to be meaningful, and updating records to a useful level of identification will require re-accessing most individual specimen records. However, entry of locality and collection event data at the time of fieldwork simplifies all subsequent uses of those data, from making specimen labels to capturing data for individual specimens.

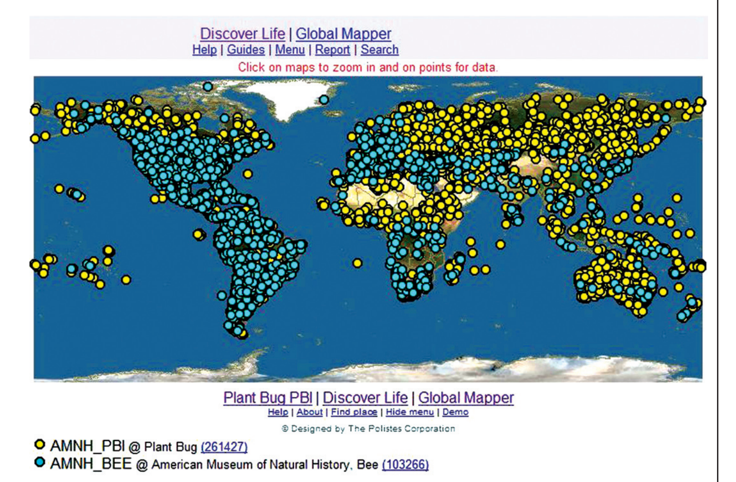


Fig. 9. Plots of all specimens from the Plant Bug PBI project and the AMNH Bee Database Project, using www.discoverlife.org. The map can be enlarged to reveal more geographic detail and additional points; all points reveal individual specimen information. Because of the large numbers of records, bee data are overlaying plant bug data, particularly in the New World.

Database Specimens as part of the Loan Process. Preparing specimen loans is a logical point for capturing specimen data because it provides a record of exactly what specimens were loaned. It may place a burden on the collection management staff because much new locality data may have to be entered. Such data could be handed to the loanee to incorporate into their own database project. As an alternative, loans can be databased upon return (from a reviser), with the benefit of more accurate identifications, but years may have passed and the precise record of what specimens were loaned will not be available, nor will existing specimen information be available for scrutiny during the interim.

Retrospective Data Capture. Because of the volume of specimens involved and lack of sufficient funds, broad-scale application of this approach is out of reach for most entomological collections, at the moment, although this situation may change if specimen databasing becomes a high priority for global, national, or regional funding agencies. Accurate identification is a prerequisite for truly useful data. Sound decision-making dictates that data capture should proceed in those groups with the highest confidence in existing identifications, where identification can be reviewed by a specialist before data capture begins, or where the data captured are of highest value for other reasons, such as comprehensive taxonomic or geographic coverage.

Accuracy and Consistency of Data Capture

Experience suggests that once data are in a database and georeferenced, they take on a life of their own because they are highly structured and have the appearance of accuracy. The more records there are, the more authoritative they may appear to be. When the data are truly accurate, they become a valuable part of a broader communal effort. If the data are not accurate, their value is illusory or positively misleading.

No matter how many control measures are instituted through the use of well-designed forms, some aspects of data capture will not be optimally consistent and therefore be the source of potential frustration. High on that list will be the names of collectors, which are often tailored to fit the available space on an insect label; consequently, the name of a single collector is often rendered in multiple ways. Localities probably rank second on this list. Once again, no single set of rules can assure uniformity, in part because data on labels are frequently rendered in varied and confusing ways, even when they are unambiguously decipherable. Good quality georeferencing should, however, allow nearly identical localities to appear in almost identical positions when the data are plotted on a map. In the end, acceptance of a certain amount of variability, even in the highly structured database environment, is the only approach that will ensure continued sanity.

Whereas some degree of variation in names of collectors and localities is tolerable, lack of enforcement of uniformity for taxon names can have a highly deleterious effect. Even here, there is no panacea for this problem because we are a long way from having an authority file of all insect (or arthropod) names. Current thinking suggests that this function will be assumed through the use of Global Unique Identifiers (GUIDs) such as those associated with all names in the Integrated Taxonomic Information System (ITIS 2009).

If the work of specimen data capture is to proceed with maximal accuracy, lists of taxon names must be under constant scrutiny by specialists. This scrutiny exists apart from the need for accurate identification. For the Plant Bug PBI project, a comprehensive list of

taxon names was bulk-loaded from the on-line Systematic Catalog of Miridae (Schuh 2002–2008), so these names never needed to be typed (or be mistyped) during routine data entry. Correcting those names and entering higher taxon names proceeds through Edit Mode. As is the case with latitude/longitude data, our collaboration with www.discoverlife.org has helped to ensure the continued accuracy of our taxonomy through the automated comparison of names we deliver with those derived from up-to-date authoritative sources available to DiscoverLife.

Database Evolution: Incorporating Multiple Users and Taxa

The Plant Bug PBI project has captured data for more than 265,000 specimens from about 50,000 localities during a 6-year period. It has also facilitated the preparation of numerous revisions and monographic works (e.g., Schuh 2006, Cassis 2008, Schaffner and Schwartz 2008). Although the PBI database began as a taxon-focused project documenting diversity in two subfamilies of plant bugs, it has been adapted to a broader range of projects spanning the United States and the world. That adaptation has involved the creation of separate interfaces [screens] tailored to the needs of the following taxon-specific projects: NSF-REVSYS: Systematics of the Scorpion family Vaejovidae; NSF-funded Global Survey and Inventory of Solifugae (http://www.solpugid.com/Database.htm); spider family Oonopidae PBI; the AMNH Bee Database Project and NSF-DBI: Collaborative Research: Collaborative Databasing of North American Bee Collections Within a Global Informatics Network, and NSF-BRC Wasp Nest conservation (AMNH 2009). All of these projects obtained substantial funding in part because each incorporated a demonstrably effective approach to specimen databasing in their proposal to the NSF. Although every project has achieved results beyond expectation, we would like to comment on two of the projects in particular.

Private funding supported the initial phases of the AMNH Bee Database Project initiated by John S. Ascher and Jerome G. Rozen, Jr. The data for this project were a natural complement to the Plant Bug PBI project because many specimen labels include information on host plants, a feature that was already available in the database. Not only has the AMNH Bee Database Project (http://research.amnh.org/iz/bee-database-project) captured data for more than 110,600 AMNH specimens, it also developed collaborations with other institutions including the University of Connecticut and Rutgers University to capture data for more than 20,000 specimens in their collections, with particular enrichment of data for the northeastern U.S. fauna. As with the plant bug project, the bee project uploaded a comprehensive list of world species to the specimen database in bulk. The source authority file was the DiscoverLife Bee Species Guide (Ascher and Pickering 2010), an updated version of the ITIS World Bee Checklist (Ascher et al. 2008). Most recently, the AMNH Bee Database project has been leveraged into a $1.4 million Biological Research Collections award from the NSF to support a multi-institutional bee data capture project starting in June 2010. The PBI database will play a central role in that effort and has allowed data capture to begin at additional institutions such as Cornell University and the Bohart Museum of Entomology, University of California, Davis, with minimal start-up time or costs.

Spiders of the family Oonopidae are the focus of the NSF-PBI project (http://research.amnh.org/oonopidae/) led by principal investigator Norman Platnick of the American Museum of Natural History. Spider specimens are preserved in alcohol as a "lot," often with multiple specimens of the same species from the same collecting event in the same vial. To accommodate this reality, we created a "lot-based" capability for the entry of specimen data; matrix code identifiers are associated with all the specimens in a vial rather than requiring the unique identification of individual specimens. The host data capture panel of the data-input screen is not shown to Oonopidae project users, simplifying the overall appearance of the interface. As with the AMNH Bee Database Project, very little additional software development was required to achieve rapid and substantial results from a worldwide team of 45 investigators (e.g., Platnick and Dupérré 2010).

Delivering Data over the Internet

Much of what we have discussed addresses use of the database by those who have login permissions. Nonetheless, our expectation is that most data will be delivered over the Internet for broader use in the community of systematists, biogeographers, ecologists, and conservation biologists, and increasingly the public, especially for charismatic groups of great interest and importance such as bee pollinators. We have addressed this issue in three ways.

First, we collaborated with John Pickering to deliver the entire contents of our database to www.discoverlife.org. DiscoverLife captures a copy of our data every night, keeping the information from our database current at all times. This provides effective and rapid placement of data and associated images on species pages, rapid and customized mapping, and coordination with the large set of botanical records also served through www.discoverlife.org, which are available for mapping, together with their associated insects.

Collaboration with DiscoverLife.org allowed us to meet project deadlines and limited expenditures on software development. Discover Life produces species pages and maps in real time (Fig. 9), and at the same time allows all users of this biodiversity portal to retain control over their own data. DiscoverLife received a large amount of high-quality data and responded by enhancing its interface to deliver improved mapping and display of information on parasite–host relationships, integrating data from all of its data providers.

Second, we deliver data through our own AMNH "species page" [http://research.amnh.org/pbi/heteropteraspeciespage]. This page acquires data from the database in real time. The assembled page includes the taxon name, images of the taxon, hosts and their images, an on-the-fly map, and detailed specimen records, the last of which can be sorted on various criteria for ease of comprehension.

Third, we have the means to make our data available through the Global Biodiversity Information Facility (GBIF). This can be achieved either directly (AMNH–GBIF) or indirectly (AMNH–DiscoverLife–GBIF).

Database Administration, Software Choice and Maintenance, and Cost

Administration. During the 5 years of PBI funding, one project participant, Sheridan Hewson-Smith, served as our database administrator, a responsibility that has now been transferred to the Division of Invertebrate Zoology. She served as the sole contact for access (administration of login permission), answered questions about usage, and collated recommendations for incremental improvements. Database users appreciate having (and often require) a contact person who can respond to their queries. Therefore, even though we worked from the beginning to allow for distributed data entry, we always focused on centralized management. During the development phase, Ms. Hewson-Smith played a crucial role in

administering software development contracts. As the usage of the database has grown, we have shifted administrative aspects of project-specific activities to lead project personnel, while at the same time maintaining overall coordination within the AMNH Division of Invertebrate Zoology.

Software Choice, Maintenance, and IT Support. The PBI project chose to use MySQL relational database software as the platform for its application development. MySQL began as a multithreaded Web-service database. Over time it has been enhanced with a greater range of functionality, such that the 5.X version now being used makes it possible to perform tasks associated with data entry, data control, and data delivery (MySQL 2010).

MySQL is also highly scalable. This means that the size of the database is essentially unlimited. This is relevant because even though a data capture effort may start out small, many projects have hit the wall when building applications in Microsoft Access, where the limit of about 500,000 records precludes use of the application for a larger effort. MySQL also allows for field sizes on the order of 4 Mb, so that large amounts of text can be stored in a single field, unlike the 256-byte restriction typical in programs like Access. These and other attributes make it possible to develop an enterprise-level database application with a long useful lifetime without the costs associated with the use of commercial software.

PBI Web pages are written in PHP. This widely used software language, also open source, is closely integrated in functionality with MySQL. The PBI project team chose to rely on Web servers at the American Museum of Natural History to serve the needs of all participants. An important aspect of this reliance is the security provided by automated enterprise-level backup procedures.

Regardless of the cost of development, the application must be maintained. This is an issue of institutional commitment. The essential requirements are as follows: first, the software must reside on a reliable server, and the physical server and its operating system software must be maintained and available at all times. Second, the application software itself must be maintained by qualified personnel. This is best accomplished at an institutional level. If this activity can be part of a larger software maintenance program, the apportioned costs will be limited. Third, the application software and the accumulating data must be backed up regularly. It takes very little time before an investment in specimen records moves into the hundreds of thousands of dollars. Thus, a data backup and archiving procedure of unquestioned reliability is essential.

Costs. Even though the software packages used in the PBI project database are essentially free, the programming is not. Nonetheless, a large community of programmers is familiar with the development of Web-based applications using MySQL and PHP. The cost of such programming is usually less than would be commanded by Oracle programmers, for example. In the end, the PBI project spent about $70,000 on application development.

Conclusions

Our efforts at creating the PBI database have led to several salient conclusions. First, all collaborators have derived great benefits from the use of a shared, Web-based collaborative database. The cost of such an approach is that it requires ongoing availability of a contact person to mediate problems and coordinate software modifications. On the benefits side, all collaborating projects enjoy the benefits of joint software and server maintenance and pooled georeferencing, and at the same time avoid the costs and problems associated with maintaining separate database installations. The PBI project was able to capture almost 80,000 specimen records for 1,300 species from the Russian Zoological Institute in St. Petersburg. Making these data available from the single most comprehensive Palearctic collection is a feat that would have been difficult to accomplish by any means other than Web-based collaboration, and it has revealed details of the published literature otherwise comprehensible only to speakers of Russian and Ukrainian. We believe this is also an example of the value of capturing transformed (as opposed to verbatim) data, in this case either translated or transliterated.

Collaboration has also facilitated our interactions with Web aggregators such as DiscoverLife because we could deliver large amounts of data from a single source, formatted in a manner amenable to mapping and associated error-checking of latitude/longitude data and taxon names. Multi-institutional collaboration has not constrained the issue of data ownership. Rather, through the use of collection-specific acronyms, all data are directly identifiable as to their source and ownership and can easily be repatriated as appropriate or necessary.

Second, curators, collection managers, and (more broadly) systematic entomologists traditionally have not incorporated unique specimen identifiers into their procedures, in distinct contrast to the practices in vertebrate zoology. In the digital age, the move to such an approach is not only valuable, it might be viewed as inevitable, and it is perhaps the only way to demonstrate the full value of collections to the public, government agencies, and other potential funders. Most PBI project participants have easily incorporated unique specimen identifiers and the use of specimen-database technology into their revisionary project workflows. Even though this has required the alteration of traditional work habits, it has yielded great benefits. These benefits include the improved accuracy and uniformity of specimen labeling, the minimization of data-capture steps, the direct assembly of information on specimens examined, the improved capability to assemble and archive voucher specimens (including those for phylogenetic and DNA-barcoding studies), and the ability to integrate specimen information across data sets, which improves taxonomic and geographic sampling and thereby extends synthetic power.

Third, attention to elements of design, particularly in the user interface, and also the underlying data structure, facilitated accurate and rapid entry of publication quality data as well as the ability to retrieve data for the widest variety of uses. By implementing consistency of functionality and logic of layout across all forms, we have been able to bring on new users with minimal training and achieve consistency of data quality.

Fourth, we have shown that a system originally designed for plant bugs can be readily modified to accommodate other phytophagous insects such as bees, and very different arthropod taxa such as spiders and scorpions. Thus, the system should prove to be generally useful for capturing arthropod collections data, and could likely be modified to accommodate a much broader range of taxa. Although we do not advocate a one-database-fits-all-approach, we believe our PBI database experience indicates that the use of Web-based approaches and cross-institutional collaboration offers substantial benefits. In our own cases, these approaches have facilitated the processes of data capture and dissemination and reduced the time and money that would have been spent on multiple freestanding database installations. Therefore, we submit our approach as worthy of consideration when planning global, national, and regional databases for entomological collections.

Acknowledgments

We thank the other members of the Plant Bug PBI team, particularly Michael Schwartz and Gerry Cassis, for their collaboration in making our efforts a success. We offer special thanks to Nina Gregorev, Division of Anthropology, AMNH, for her dedication to creating a truly successful user interface, and Paul Flemons of the Australian Museum for his contributions to the database design. Thanks also go to Mark Breedlove, formerly of the AMNH, for implementing the database on the AMNH Web servers and maintaining the Web application; Thomas Trombone, AMNH Division of Vertebrate Zoology, and the late James S. Ashe, University of Kansas, for their contributions to portions of the logic incorporated into the final application and its ultimate effectiveness; Sacha Spector, formerly AMNH, for his help in implementing the use of matrix codes for unique specimen identifiers and other valuable input; and Lorenzo Prendini, Norman Platnick, and Jerome G. Rozen, Jr., Division of Invertebrate Zoology, AMNH, and John Pickering, www.discoverlife.org, for their support of the project.

Funding for the Plant Bug PBI project and database development was provided by NSF award DEB0316495. Additional support for database development came from the AMNH Bee Databasing Project, funded by AMNH trustee emeritus Robert G. Goelet, with continuing funding to the American Museum of Natural History from NSF award DBI0956388. The Web-based mapping tool was created by David Shorthouse from the Encyclopedia of Life staff, with funding from the NSF-PBI project for Oonopidae spiders, DEB0613754, Norman Platnick principal investigator.

References Cited

AMNH (American Museum of Natural History). 2009. http://research.amnh.org/iz/hymenoptera/collection/

Ascher, J. S., and J. Pickering. 2010. DiscoverLife Bee Species Guide and World Checklist (Hymenoptera: Apoidea: Antho. http://www.discoverlife.org/mp/20q?guide=Apoidea_species&flags=HAS.

Ascher, J. S., et al. 2008. ITIS World Bee Checklist. http://www.itis.gov/beechecklist.html

BioGeomancer. 2010. http://www.biogeomancer.org/.

Cassis, G. 2008. The *Lattinova* complex of austromirine plant bugs (Hemiptera: Heteroptera: Miridae: Orthotylinae). Proc. Entomol. Soc. Wash. 110: 845–939.

Colwell, R. K. 2009. Biota–the biodiversity database manager. http://viceroy.eeb.uconn.edu/Biota.

Forero, D. 2008. Revision and phylogenetic analysis of the *Hadronema* group (Miridae: Orthotylinae: Orthotylini), with descriptions of new genera and new species, and comments on the neotropical genus *Tupimiris*. Bull. Am. Mus. Nat. Hist. 312.

Fuzzy Gazetteer. 2010. http://isodp.fh-hof.de/fuzzyg/query/.

GBIF (Global Biodiversity Information Facility). 2010. http://www.gbif.org/.

GEOLocate. 2010. http://www.museum.tulane.edu/geolocate/default.aspx.

GNIS (Geographic Names Information System). 2010. http://geonames.usgs.gov/pls/gnispublic.

ITIS (Integrated Taxonomic Information System). 2009. http://www.itis.gov/.

Interagency Working Group on Scientific Collections (National Science and Technology Council, Committee on Science, Interagency Working Group on Scientific Collections). 2009. Scientific collections: mission-critical infrastructure of federal science agencies. Office of Science and Technology Policy, Washington, DC. http://www.whitehouse.gov/sites/default/files/sci-collections-report-2009-rev2.pdf.

MySQL. 2010. http://www.mysql.com/.

Platnick, N. I., and N. Dupérré. 2010. The goblin spider genus *Scaphiella* (Araneae, Oonopidae). Bull. Am. Mus. Nat. Hist. 332.

Schaffner, J. C., and M. D. Schwartz. 2008. Revision of the Mexican genera *Ficinus* Distant and *Jornandes* Distant with the description of 21 new species (Heteroptera; Miridae: Orthotylinae: Orthotylini). Bull. Am. Mus. Nat. Hist. 309.

Schuh, R. T. 2006. Revision, phylogenetic, biogeographic, and host analyses of the endemic western North American *Phymatopsallus* group, with the description of 9 new genera and 15 new species (Insecta: Hemiptera: Miridae: Phylinae). Bull. Am. Mus. Nat. Hist. 301.

Schuh, R. T. 2002–2008. On-line systematic catalog of plant bugs (Insecta: Heteroptera: Miridae). http://research.amnh.org/pbi/catalog/.

Schuh, R. T., and P. Pedraza. 2010. *Wallabicoris*, new genus (Hemiptera: Miridae: Phylinae: Phylini) from Australia, with the description of 37 new species and an analysis of host associations. Bull. Am. Mus. Nat. Hist. 338.

Specify 6. 2010. http://specifysoftware.org/.

Tatarnic, N. J., and G. Cassis. 2008. Revision of the plant bug genus *Coridromius* Signoret (Insecta, Heteroptera, Miridae). Bull. Am. Mus. Nat. Hist. 315.

TDWG (Taxonomic Database Working Group). 2010. http://www.bgbm.org/TDWG/acc/Referenc.htm.

Randall T. Schuh is the George Willett curator, Division of Invertebrate Zoology, American Museum of Natural History. His primary research interests are in the systematics of Heteroptera, particularly the Miridae. He can be reached at schuh@amnh.org. **Sheridan Hewson-Smith** is a native of Australia with a background in Environmental Management and Education. She served as the PBI-project technical coordinator with responsibilities in the areas of software development, georeferencing, data-quality management, and project coordination. She can be contacted at: shs.amnh@yahoo.com. **John S. Ascher**, who received his Ph.D. in entomology from Cornell University, is manager of the AMNH Bee Database Project. He can be reached at ascher@amnh.org.